Potential Risk Indexing System (P-RISK Model) Utilizing GIS to Rank Geographic Areas, Industrial Sectors, Facilities, and Other Areas of Concern

Debra L Forman (1), Amy Amina C Wilkins (2),* David West (3)
(1) Waste and Chemicals Management Division, US Environmental Protection Agency, Region III, Philadelphia, PA; (2) Office of Research and Development, National Center for Environmental Assessment, US Environmental Protection Agency, Washington, DC; (3) Office of Policy and Management, US Environmental Protection Agency, Region III, Philadelphia, PA

Abstract

The Potential Risk Indexing System (P-RISK) is a screening methodology and computer-based program that ranks areas of concern (i.e., facilities, industrial sectors, and geographic areas) according to multi-media chemical releases, chemical toxicities, and selected demographics of surrounding populations. The model uses geographic information system (GIS) technologies to display vast quantities of data, assisting users in cumulative risk analysis and other decision-making processes. P-RISK users include risk assessors and managers, US Environmental Protection Agency (EPA) program offices and Regions, state environmental departments, and other communities concerned with environmental targeting, inspection targeting, pollution prevention targeting, resource prioritization, environmental justice analysis, trend analysis, and comparative risk efforts. P-RISK operations currently include five steps. First, release data are retrieved from the EPA's Toxics Release Inventory System, the Aerometric Information Retrieval System, and the Permits Compliance System for reported chemical emissions to air, land, and water. Second, toxicity values are obtained from the Integrated Risk Information System and the Health Effects Assessment Summary for oral carcinogenic and non-carcinogenic effects. A dose is then calculated and weighted to create an index of relative toxicity scores. Third, a visual GIS data layer showing color-coded, indexed areas (based on an 8-mile by 8-mile grid system) ranging from high-release and high-toxicity combinations to low-release and low-toxicity combinations is created and made available for viewing and printing. Fourth, in keeping with EPA environmental justice and children's health protection guidelines, a second GIS color-coded, indexed data layer is generated, using US Census data, to ascertain income and minority status as well as other factors that may influence the relative vulnerability of subpopulations. This approach does not yet imply exposure, which must be assessed using intake parameters adjusted for specific subpopulations. Presently, the vulnerability index created in step 4 characterizes potentially exposed populations, highlighting those that may be more vulnerable. In this way, the screening can include population characteristics without using broad assumptions about exposure conditions. Finally, the chemical/facility index (created in step 3) and the vulnerability index can then be overlaid to match incidence of high-toxicity, large-release combinations with areas having relatively high percentages of vulnerable populations.

* Amy Amina Wilkins, US Environmental Protection Agency, Office of Research and Development, National Center for Environmental Assessment, 401 M St., SW, Mail Code 8623-D, Washington, DC 20460 USA; (p) 202-564-3256; (f) 202-565-0078; E-mail: Wilkins.Amina@epamail.epa.gov

Keywords: facility ranking or index, demographic populations, environmental justice

Introduction: The Potential Risk Indexing System

The Potential Risk Indexing System (P-RISK) is a computer-based screening model that ranks facilities, industrial sectors, or geographic areas according to data on multi-media chemical releases and chemical toxicities, population demographics, and other spatial features. The P-RISK is structured to enable the user to consolidate all available data, ranging from ambient air and water status to watershed health to potential human health risk, in one user-friendly product.

By using geographic information system (GIS) software to display vast quantities of data, the P-RISK assists users in cumulative risk analysis, broadening individual programmatic criteria for targeting enforcement actions in stressed areas, or identifying improved environmental protection in other areas. Expected users of the P-RISK include US Environmental Protection Agency (EPA) risk assessors, EPA risk managers, and state and local environmental regulatory agencies concerned with inspection targeting, pollution prevention targeting, resource prioritization, environmental justice analysis, trend analysis, and comparative risk efforts. Communities that use environmental information in their decision-making may also use the P-RISK.

The P-RISK has resulted from the consolidation of two independent development efforts, the Chemical Indexing System (CIS) (1,2), prepared by EPA Region III, and the Risk-Based Enforcement Strategy (RBES) (3,4), prepared by EPA's National Center for Environmental Assessment at the Office of Research and Development. As part of the consolidated development process, the P-RISK Workgroup, consisting of EPA staff from headquarters and the Regions, has been convened to provide critical input and help develop the P-RISK model as it currently exists. This peer-review package includes input provided by the workgroup during a six-month period from June to November 1998 (5).

Conceptual Model

The P-RISK is intended to address several complex questions pertaining to potential risk and exposure to toxics. It is designed to answer a range of questions from a variety of users using the familiar platform of a personal computer (PC). To do this, the P-RISK takes the following three-step approach:

1. Retrieval of data from EPA databases.
2. Manipulation of these data to calculate a set of indices executed within a series of independent modules.
3. Use of a GIS interface to integrate the data and test scenarios for potential risk, exposure, and compliance.

The modules are arranged to reflect the 1983 National Academy of Sciences risk paradigm (6) and are tooled to permit independent execution of each module. This structure maximizes flexibility for the user and enables a description of uncertainty at each stage of the analysis.

The first step, data retrieval, uses a SAS program (SAS Institute, Inc., Cary, NC) to gather chemical release information from a variety of EPA databases. Currently, data for reported chemical releases to air, land, and water are retrieved from the Toxics Release Inventory (TRI) (7) and the Permit Compliance System (PCS) (8) housed in the ENVI-ROFACTS data tables (9) on EPA's Valley Unix machine in Research Triangle Park, North Carolina. Data from the Aerometric Information Retrieval System (10), the Resource Conservation and Recovery Information System (11), and the Biennial Reporting System (12) will be added to this system.

The second step in the process generates a suite of indices, described below. (See Figure 1.) Briefly, the Chronic Index, contained in the P-RISK system's P-RISK Module, is based on the volume and toxicity of a chemical release and characterizes a chemical, facility, or Standard Industrial Classification (SIC) of interest to the user. The Vulnerability Index is divided into two components, census and disease. The Vulnerability Index: Census, also contained in the P-RISK Module, describes demographic characteristics that may render a subpopulation more likely than the general population to be harmed by toxic chemical exposures. The Vulnerability Index: Disease describes existing disease incidences and is included in the Exposure Module, along with the Exposure Index. The Compliance Index, contained in the Compliance Module, provides a statutory component to describe current and historical conformance with environmental laws.

To permit the greatest level of flexibility, the model has been developed to calculate each index independently, allowing the user to refine the analysis in a stepwise fashion and gain an appreciation for the data gaps and uncertainties at each stage of the analysis. The P-RISK's modules, some of which are currently under development, are briefly described below.
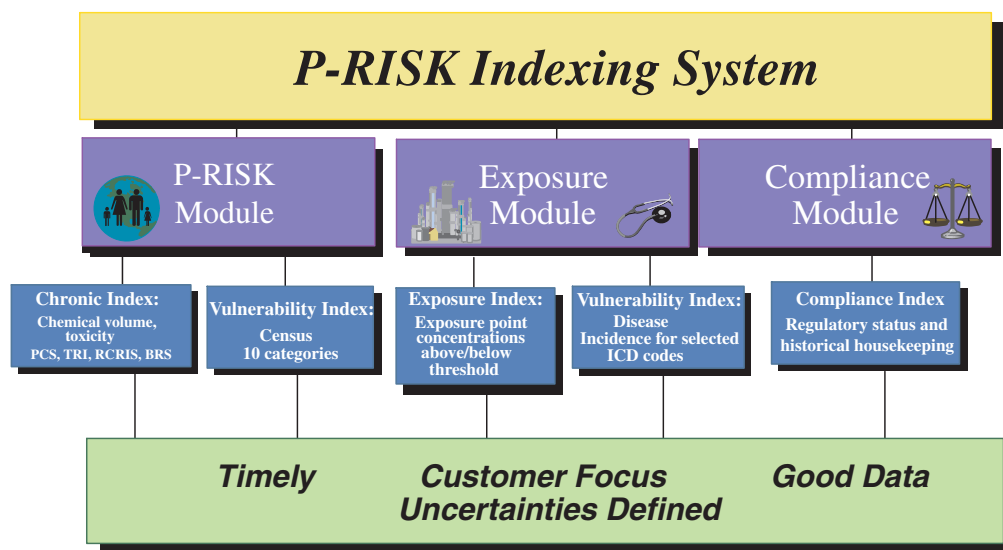

Figure 1 P-RISK structure.

The P-RISK Module has two components:

- The *Chronic Index*, a relative rank using scores derived from the reported volume of each chemical released at a site along with its associated toxicity. Individual Chronic Indices can be ranked or summed according to a given selection of chemical, facility, or SIC. The user has the option of generating these data for any or all of the 50 US states or the 10 EPA Regions. The user also has the option of
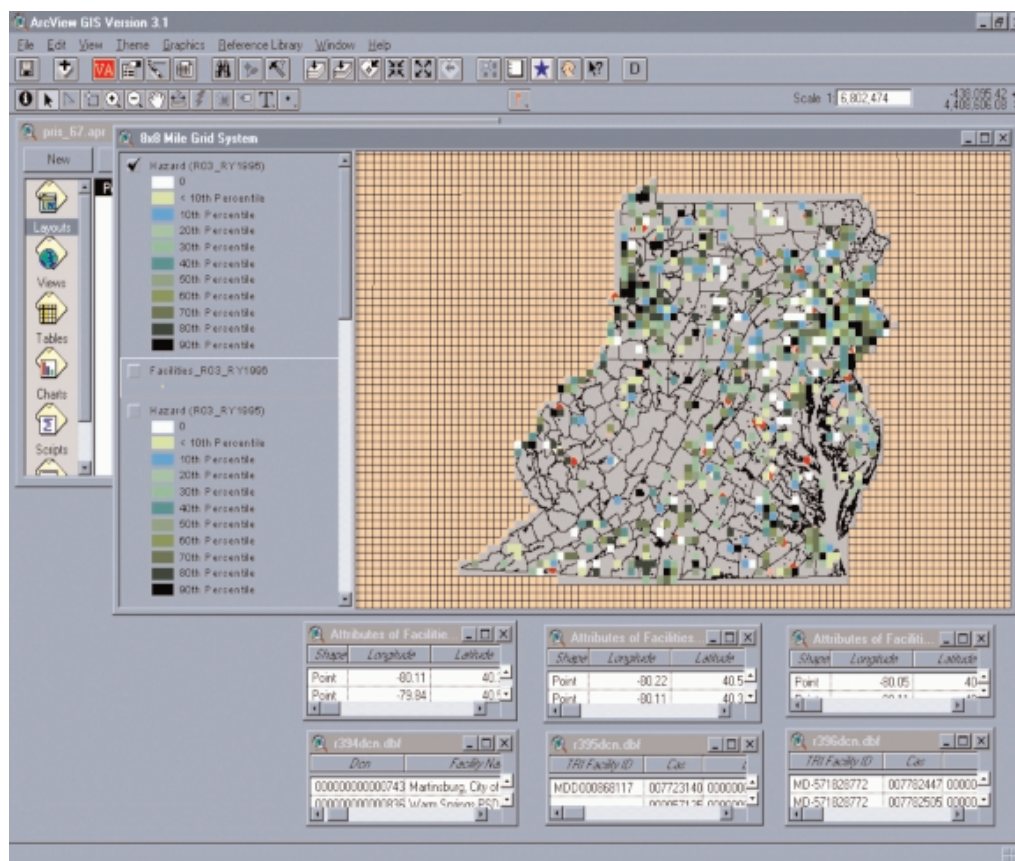

Figure 2   P-RISK interface, Chronic Index view.

creating these data as Lotus (Lotus Development Corporation, Cambridge, MA) spreadsheets or dBASE III (Ashton-Tate Corporation, Torrance, CA) files, which are stored on the user's PC and further processed to create a GIS data layer for use with the ArcView (ESRI, Redlands, CA) GIS software (Figure 2).

- The *Vulnerability Index: Census*, which provides a description of socioeconomic and demographic characteristics that may render a subpopulation more likely than the general population to be harmed by exposure to toxic chemicals. Given that "vulnerability" can mean different things to different people, for this application, the definition of vulnerability is consistent with recommendations of the

P-RISK Workgroup and in keeping with the environmental justice guidelines developed by EPA. Namely, "vulnerability" encompasses both biologically sensitive populations and potentially highly exposed populations. In general, this index is intended to highlight those populations that may be more vulnerable. In keeping with EPA environmental justice (13) and children's health protection guidelines (14), US Census data are available on CD-ROM for all 50 states. The current project uses Census data for the District of Columbia and the five states in Region III: Delaware, Maryland, Virginia, West Virginia, and Pennsylvania. The project contains Census data for the following nine fields: minority status, poverty status, age, pregnancy, female head of household with children and no husband present, educational attainment, unemployment, and age of home.

The first module, P-RISK, purposely does not account for exposure, which must be assessed using intake parameters adjusted for specific populations. Moreover, it is important to note that the term "vulnerable" does not imply causality or a mathematical relationship between a demographic characteristic and a hazard. Rather, the data are intended as an illustration, demonstrating the potential for risk and alerting the user that further analysis may be warranted.

The Exposure Module has two components:

- The *Vulnerability Index: Disease*, which includes a description of disease incidence using county-level data provided by the federal Centers for Disease Control and Prevention. A finite number of International Classification of Diseases (ICD) codes were selected based on the critical effects of priority compounds. Professional judgement should be an integral part of the ICD selection process. Clearly, the manifestation of disease can have many causal factors, including diet, lifestyle, smoking, and environmental insults. Disease occurrence cannot prove a causal relationship with nearby sources; therefore, the purpose of this index is to demonstrate the current disease burden for a selected area and provide a public health context for the P-RISK Module.

- The *Exposure Index,* which provides a method for evaluating potential exposures. This index is the most complex and possesses the greatest uncertainty. Several models are currently under consideration for inclusion in the P-RISK, including RBES (3,4), the Ecological Sensitivity Targeting and Assessment Tool (15), the Cumulative Exposure Project (16), and the Total Risk Integrated Methodology (17). For each of these models, characterization of the uncertainty in the resultant estimates is a key feature.

The Compliance Module has one component:

- The *Compliance Index*, which provides a compliance component for targeting purposes. Several existing compliance rating methodologies are currently under consideration for inclusion in P-RISK, including the Site Index Database (18) rating factors and the Sector Facility Indexing (19) approach to compliance. The Compliance Index takes advantage of compliance history both in terms of issued complaints and the complaint patterns derived from the Emergency Response Notification System database (20) and the Integrated Data for Enforcement Analysis system (21,22). The Compliance Index can serve to highlight areas with a high potential for non-notifiers by focusing on geographic or industry sectors

with large numbers of complaints or poor compliance history. At the same time, the Compliance Index enables the user to identify facilities that have had exceptional environmental and compliance histories and could serve as models for the regulated community.

The final step of the P-RISK model allows users to view and interact with the information contained in the indices, including pertinent geographical datasets that may be available on the user's home system. Using ArcView, the user creates maps, queries the data, and creates new themes on demand according to user-specified criteria. One specific coverage included in the first module of P-RISK is an 8-mile by 8-mile grid of the contiguous United States displaying areas of relatively high volume and high toxicity for TRI and PCS industrial releases for the reporting year of interest (i.e., the Chronic Index). The same module also includes demographic coverages at the census block level to display counts for each of the nine different demographic categories (i.e., the Vulnerability Index: Census).

The project also contains a reference library that gives access to several ecosystem and municipal coverages, including ambient air monitoring data, stream alkalinity, a relative ranking of watershed health, public lands, cities, and zip codes. Each of the coverages in the system is currently available for Region III, but may be revised to include coverages available on the user's home system. Each coverage is documented in the reference library, including a description, a justification for use, and a source citation. With live access to all data in one place, the user can view and query the coverages in a spatial environment, matching incidences of large-volume, highly toxic releases with areas of either complaints, high percentage of vulnerable populations, or vulnerable ecosystems.

ArcView GIS Project

The P-RISK model is delivered in two parts: (1) a user-friendly, interactive ArcView GIS project that can create maps, query data, and create new themes on demand according to user-specified criteria, and (2) an automated SAS program that can be customized to user specifications for risk level, area of concern, and type of output. The objective of P-RISK is to provide the user with real-time access to the data and the discretion to overlay, view, and query the different indices generated in the modules.

Data Analysis

The following steps are taken to produce the data layers or themes for the GIS maps. All coverages in the ArcView project are projected in albers equal-area conic with spheroid GRS 80, central meridian equal to -79, first standard parallel equal to 37, second standard parallel equal to 41, and latitude of projection's origin equal to zero. There is no false easting or northing; both are set to zero. The datum is the North American Datum of 1983, or NAD 83.[1]

[1] For more information on NAD 83, see the Geodetic Glossary (Rockville, MD: National Geodetic Survey, National Ocean Service, National Oceanic and Atmospheric Administration, September 1986) or see questions 2 and 3 at http://www.ngs.noaa.gov/faq.shtml.

Chronic Index Coverage

SAS code is used both to retrieve data from the EPA databases and to calculate the Chronic Index. The final output is provided as four files (one each for chemical, document control number, facility, and SIC code) in either Lotus or dBASE format and is stored on the user's PC. The algorithm also contains a quality assurance check to ensure that the code is executing properly.

P-RISK allows the user the opportunity to map the Chronic Index by facility, SIC code, and chemical. The output files, stored on the user's PC, are retrieved from within ArcView to create a color-coded grid expressing potential risk. This data layer can be manipulated to illustrate the rank of a chemical release in a particular grid, the facilities responsible for the toxic chemical release, and the SIC codes that are primarily responsible for the toxic release. The user has the option to create files either in Lotus or dBASE format. The user's manual provides details on how the program calculates the Chronic Index and aggregates the data by chemical, facility, and SIC code.

Census Block and Demographic Information

Census block group coverages or data layers are available from various electronic sources. For example, the Regional census block group boundary coverages can be found at the following EPA FTP site: ftp://valley.rtpnc.epa.gov. Because these electronic sites may vary from Region to Region, each EPA Regional GIS group should be contacted for specific boundary coverage locations. Once the boundary coverage has been obtained, it can be coupled with demographic information to create additional data layers. In the Region III example, the data for nine census fields (downloaded from the 1990 US Census CD-ROMs) were permanently joined to the block group boundary coverage. For each of the nine census fields, the block groups are sorted from lowest to highest and divided into ten equal percentiles. This ranking permits the user to view any percentile for any census field, such as the top 10% of the Region's census block groups for poverty.

The Fetch Button

The Fetch button is used to process the TRI/PCS facilities database. After clicking on the button, the system prompts the user for the location of the facilities database file to process. The data are then retrieved into ArcView by Fetch and run through checks to see if the proper information is contained in the file. Once the database file passes the checks, a point coverage is created from the available latitude and longitude data. If the latitude and longitude field is not populated, then a point is created using the zip code centroid. Statistics are then run on the new facility point coverage to create 10 percentiles for each of the nine chemical release categories:

- Fugitive air releases
- Stack air releases
- Water releases (containing releases from either PCS or TRI)
- Underground releases
- Land releases
- Onsite releases
- Releases from publicly owned treatment works

- Offsite releases
- Total releases and transfers

The new facility point coverage and its attached percentile database are then added to the view and a legend is created. From the new facility coverage, an 8-mile by 8-mile grid aggregation of the facility points is created. Statistics are also run in the 8-mile by 8-mile aggregated grid to produce percentiles for each of the nine chemical release categories. After adding the grid theme to the view, the system prompts the user for the location of the chemical table. Once the corresponding chemical table is found, it is retrieved into ArcView by Fetch and linked to the facility point coverage. This marks the successful completion of the program, leaving the user with a ranked, color-coded grid coverage and a ranked facility point coverage linked to individually ranked chemical releases. The P-RISK retains up to three years of processed data to enable the user to investigate recent trends in hazard.

Blue Star Query

From this data universe, the Blue Star button allows the user to select a subset of information contained in any data theme in P-RISK. For instance, the user can use the Blue Star to select the top 10% of the fugitive air combined index to identify and display those facilities releasing chemicals with the highest volumes and toxicities. The user is prompted through a series of help menus to locate the item of interest. Once the query is completed, the resultant subset is linked to both the facility and chemical tables so that the user can determine specific chemicals, their sources, amounts, rank, carcinogenic or non-carcinogenic toxicity, uncertainty, and other attributes.

Red Flag Query

The Red Flag button allows users to select individual sites from the subset created with the Blue Star. Using this feature, the user can access the row within the data table associated with the particular site. All selected sites can be moved to the top of the table by clicking on the top border of the table and then clicking the "promote" icon.

Reference Library

The Reference Library enables the user to access other available data coverages, consolidating desired information in one place. This structure facilitates place-based analyses and assists in identifying the relative contribution of several different impacts. Online documentation is also available.

References

1. US Environmental Protection Agency (EPA). 1995. *Addendum to Chemical Indexing System Part I: Chronic Index technical guidance manual.* EPA Region III, Air, Radiation and Toxics Division. EPA/903/R-93/002-a. August.

2. US Environmental Protection Agency (EPA). 1993. *Chemical Indexing System Part I: Chronic Index technical guidance manual.* EPA Region III, Air, Radiation and Toxics Division. EPA/903/R-93/002.

3. US Environmental Protection Agency (EPA). 1994. *Risk-Based Enforcement Strategy: Final report, September 30, 1994.* EPA Contract 68-D3-0013, Task Number 2-15. Prepared by Versar, Inc., Springfield, VA.

4. US Environmental Protection Agency (EPA). 1994. *Risk-Based Enforcement Strategy user's guide, September 30, 1994.* EPA Contract 68-D3-0013, Task No. 2-16. Prepared by Versar, Inc., Springfield, VA.

5. US Environmental Protection Agency (EPA). 1998. *Peer review package for the Potential Risk Indexing System (PRIS), September 30, 1998.* Report submitted to EPA by Science Applications International Corporation, McLean, VA. EPA Contract 68-D4-0098, WA IV-5.

6. US Environmental Protection Agency (EPA). 1992. *Guidelines for exposure assessment. Risk Assessment Forum.* EPA/600-Z-92-001.

7. US Environmental Protection Agency (EPA). 1997. *Toxics Release Inventory (TRI) relative risk-based environmental indicators: Methodology, June 1997.* Washington, DC: Office of Pollution Prevention and Toxics, US Environmental Protection Agency. EPA/740-R-97-002.

8. US Environmental Protection Agency (EPA). 1997. *Permit Compliance System (PCS) enforcement action national file* (on diskette). Washington, DC: Office of Wastewater Enforcement and Compliance, US Environmental Protection Agency. NTIS PB97-502314.

9. US Environmental Protection Agency (EPA). 1999. *ENVIROFACTS data warehouse and applications.* Office of Environmental Information, US Environmental Protection Agency. http://www.epa.gov/enviro/.

10. US Environmental Protection Agency (EPA). 1994. *Aerometric Information Retrieval System (AIRS) user's guide.* Washington, DC: National Air Data Branch, Technical Support Division, Office of Air Quality Planning and Standards, US Environmental Protection Agency. EPA-454/B-94-007.

11. US Environmental Protection Agency (EPA). 1998. *Resource Conservation and Recovery Information System (RCRIS) merged database administrator guide: v.7.0.0, February 1998.* Washington, DC: Office of Solid Waste, US Environmental Protection Agency. SCD-0055-033-RL-7051.

12. US Environmental Protection Agency (EPA). 1995. *Biennial Reporting System (BRS) data: Generation and management of hazardous waste, 1995 (final)* (on CD-ROM). NTIS PB98-500077. Washington, DC: Office of Solid Waste, US Environmental Protection Agency.

13. White House. 1994. *Federal actions to address environmental justice in minority populations and low-income populations: Executive Order 12898.* President William J. Clinton. Washington, DC. Executive Order 12898. February 11, 1994.

14. White House. 1997. *Protection of children from environmental health risks and safety risks: Executive Order 13045.* President William J. Clinton. Washington, DC. April 21, 1997.

15. US Environmental Protection Agency (EPA). 1999. *Ecological Sensitivity Targeting and Assessment Tool (ESTAT).* Science Applications International Corporation. http://www.epa.gov/envirofw/html/factsheets/estat.html.

16. US Environmental Protection Agency (EPA). 1999. *Cumulative Exposure Project (CEP).* Office of Policy, US Environmental Protection Agency. http://www.epa.gov/CumulativeExposure/home.htm.

17. US Environmental Protection Agency (EPA). 1999. *Total Risk Integrated Methodology (TRIM) technical support document: External review draft.* Research Triangle Park, NC: Office of Air Quality Planning and Standards, US Environmental Protection Agency. EPA-453/D-99-001. November.

18. Williams S. 1998. Personal communications with Stephen Williams, Delaware Department of Natural Resources and Environmental Control. Dover, DE. June.

19. US Environmental Protection Agency (EPA). 1997. *Science Advisory Board report: Review of the Sector Facility Indexing Project (SFIP).* Washington, DC: EPA Science Advisory Board. EPA-SAB-EEC-97-012.

20. US Environmental Protection Agency (EPA). 1998. *Emergency Response Notification System (ERNS), 1996* (on CD-ROM). Washington, DC: Office of Emergency and Remedial Response, US Environmental Protection Agency. NTIS PB98-500663.

21. US Environmental Protection Agency (EPA). 1996. *Integrated Data for Enforcement Analysis (IDEA) system.* Office of Enforcement and Compliance Assurance, US Environmental Protection Agency. EPA/300/B-95/004D.

22. US Environmental Protection Agency (EPA). 1996. *IDEA basic training course: Student booklet, Integrated Data for Enforcement Analysis, March 1996.* National Technical Information Service (NTIS). PB96-780507.